Cureus Journal of Computer Science

Open Access Original Article

Privacy-Preserving and Information-Sharing Mechanisms Using Marine Fish Optimization for Big Data Applications

Received 09/18/2025
Review began 09/27/2025
Review ended 10/29/2025
Published 11/08/2025

DOI:
https://doi.org/10.7759/s44389-025-10497-y

Ravi Mante [1] , Prashant N. Chatur [1] , Milind B. Waghmare [1]

1. Computer Science and Engineering, Government College of Engineering, Amravati, IND

Corresponding author: Ravi Mante, mante.ravi@gmail.com

Abstract

Big data refers to extensive, intricate, and expanding datasets that originate from various independent sources. In recent years, big data has attracted significant attention as it has the potential to generate new insights that foster technological innovation and drive economic development. Consequently, researchers have started to focus their efforts on processing big data to address the challenges associated with its high volume, velocity, and variety, commonly referred to as the "3Vs." In addition to the 3V challenges, new privacy and security concerns are emerging within the realm of big data. Techniques for data mining are utilized to extract beneficial knowledge from large amounts of data, also referred to as big data. Many data mining methods exist that can assist in finding valuable knowledge from complex relationships within big data, as well as from dynamically changing volumes, especially when the amount of data tends to shift rapidly.

Preserving the privacy and integrity of private and public data shared by different entities in a big data framework is a major concern. This research proposes a model for preserving privacy in data sharing within the big data framework, known as Marine Fish Optimization (MFO)-Privacy Preservation. MFO is a hybrid approach that combines the behaviors of both marine predators and jellyfish. The optimization within this model includes a privacy key optimizer based on the behavioral principles of marine predators and jellyfish. The algorithm generates keys to modify data clusters to improve data privacy while maintaining data integrity and usability. This hybrid approach (MFO) addresses complex optimization challenges. Based on the results, MFO achieved a data privacy rating of 1.869, with a decryption time of 0.02249 and an encryption time of 0.329.

Categories: Big Data and Analytics, Security and Privacy, Data Mining
Keywords: privacy preservation, information sharing, privacy-preserving, mfo, integrity

Introduction

Big data is related to large, complex, and increasing datasets that have many autonomous sources. Big data has recently attracted interest due to the possibility of acquiring new insights from the data for economic growth and associated technical innovation, and there are now many research activities associated with big data processing because of the size, speed, and variety of data (referred to as "3V") challenges [1,2]. Along with 3V challenges, there are newly arising privacy and security challenges in big data [3-6]. In a big data framework, data is distributed. Chen et al. [7] have introduced a learning approach to Bayesian networks from distributed odd data. Large datasets can be mined for valuable information. Numerous mining methods have been created to extract useful information from large datasets, which frequently contain complex relationships and dynamically changing ratios [8,9]. Therefore, an effective theoretical and technical framework is required to support data stream mining.

The word "big data" has come from the fact that users produce huge amounts of data daily. About 2.5 quintillion bytes of data is created daily, and today, 90% of the world's data has been created in the last two years [10,11]. After the arrival of information technology in the early 19th century, the ability of users to generate data production was never so vast and powerful. For example, on 4 October 2012, the first presidential discussion between President Barack Obama and Governor Mitt Romney generated more than 10 million tweets in two hours [12]. In numerous tweets, specific moments - such as those involving Medicare and the voucher program - were the most discussed during the discussion. These online conversations serve as a real-time barometer of public interest, which underlines the most interesting issues to the audience. Unlike traditional media platforms like radio or television, which provides unilateral dialogue, social media enables dynamic, two-way communication, which makes it a more attractive and immediate way to measure public emotions. This real-time feedback mechanism provides valuable insights for policymakers, media analysts, and researchers, which can better understand and respond to changing public anxiety. Another instance is Flickr, a publicly accessible photo-sharing platform, which received an average of 1.8 million photos daily from February to March 2012 [13]. Assuming that each image is 2 megabytes in size, this indicates that a daily storage capacity of 3.6 terabytes is necessary. As the saying goes, "A picture is worth a thousand words." The billions of photos on Flickr serve as a valuable resource for examining human society, social initiatives, public events, disasters,

How to cite this article
Mante R, Chatur P N, Waghmare M B (November 08, 2025) Privacy-Preserving and Information-Sharing Mechanisms Using Marine Fish Optimization for Big Data Applications . Cureus J Comput Sci 2 : es44389-025-10497-y. DOI https://doi.org/10.7759/s44389-025-10497-y

and much more for users, provided that users possess the ability to utilize such vast amounts of data.

The examples above highlight the rapid expansion of big data applications, where data generation has surged to levels that exceed the handling capabilities such as capturing, managing, and processing of traditional software tools within acceptable time limits. A fundamental challenge in big data is to analyze massive datasets and extract useful insights or knowledge to guide future decisions. Often, this extraction must be performed efficiently and nearly in real time, as storing all incoming data is generally not feasible.

Challenges of big data

To effectively manage massive datasets in intelligent educational database systems, it is essential to assess the vast volumes of data and address the unique features identified by the HACE (Heterogeneity, Autonomy, and Complexity at an Exponential scale) theorem. As illustrated in Figure *1* [14], the conceptual framework for processing big data includes three tiers: data access and computation (Tier I), data privacy and domain-specific knowledge (Tier II), and big data mining algorithms (Tier III). Research challenges in big data mining can systematically be organized in three-level frameworks; each level focuses on the specific level of abstract and technical complexity. This layered model helps to explain how challenges develop from infrastructure and computer problems to algorithm intelligence at the highest level of algorithm [14].

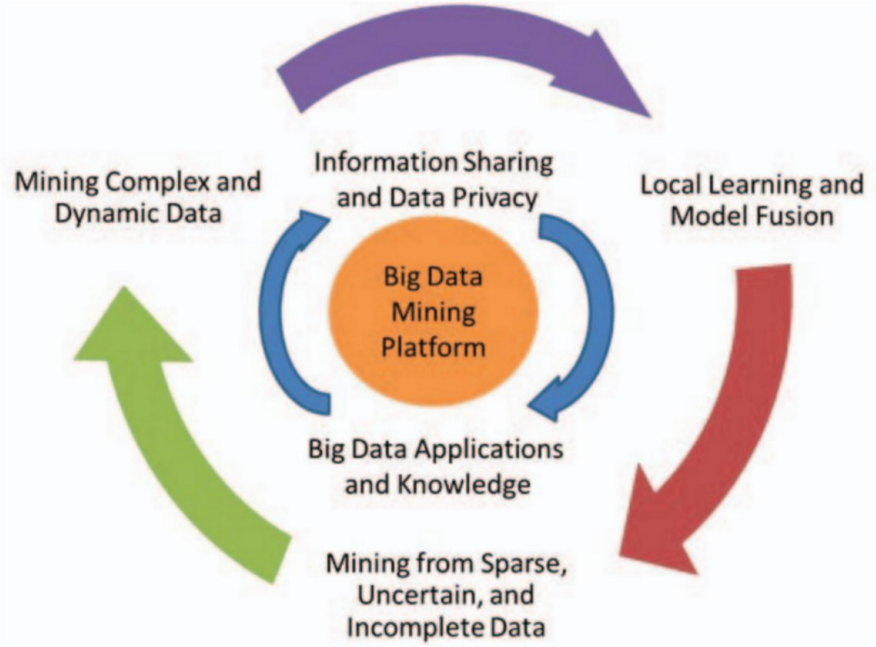

FIGURE 1: A Big Data Processing Framework

The challenges in Tier I are concentrated around the data access and arithmetic processes. Large data is often stored in many places, and their proportion continues to increase, so effective computer platforms have to consider the distributed large amount of storage. For example, the traditional data-linked algorithm requires loading all the data in the main memory, but it has become a clear technical barrier for large data because moving data in different places is expensive (for example, due to more network traffic and other input/output costs), even though a very large main memory is available for all data for calculations [14].

Tier II challenges emphasize the economic aspects and domain-specific expertise required for various big data applications. This knowledge can enhance the data mining process while also interacting with the technical issues related to data access (Tier I) and the development of mining algorithms (Tier III). For example, data manufacturers and data sharing systems in data manufacturers that depend on different domain applications may vary significantly. Sensor network data may not be disappointed for applications like water quality maintenance, but to issue and share the location of mobile users for most applications, it is not clearly acceptable. In addition to all of the above privacy issues, an app domain can also yield more information to inform or guide the app to the big data mining algorithm design. In market basket transaction data, for example, all transactions are viewed as independent, and the aim of the data is generally represented by searching for very highly correlated objects, perhaps under different temporal and/or spatial constraints. On the contrary, in social networks, users are connected through links and share dendritic structures that represent their interdependencies. The knowledge is then represented by

user communities or groups, including their leaders and the social outcomes of each group. Overall, it is crucial to tap into around four levels of semantics and application knowledge to target lower levels for data access and later design the high-level mining algorithm [14].

Tier III landscape employs a three-stage cycle. In the first stage, a fusion approach is used to preprocess incomplete, uncertain, multisource, heterogeneous, and sparse data. The second stage involves mining dynamic and complex data after preprocessing. The third stage focuses on testing the performance of global knowledge derived from local learning and model fusion. In this stage, relevant knowledge is also derived, and information related to the preprocessing stage is retrieved. Based on feedback, models and parameters can then be adjusted accordingly. At all stages, information sharing is not only essential for ensuring a smooth transition between stages but also reflects the fundamental purpose of data mining within big data processes [14].

However, this exponential growth in data and finding interesting knowledge from big data is linked to an equally important concern, the need to protect personal privacy. As data becomes increasingly ubiquitous, the risk of personal privacy being violated through inadvertent data disclosure and malicious attacks becomes increasingly serious [15]. The challenge is the vulnerability of quasi-identifiers, those tenuous but potentially revealing features that, when combined, could reveal a person's identity. In the big data space, attribute disclosure and identity disclosure attacks are occurring at a large scale in the background, obscuring the responsible use of data. In response to this important concern, our research embarks on a transformative journey. Here we present a pioneering privacy protection model explicitly tailored for big data applications [16,17].

For big data applications, privacy-protecting models that use optimal key generation approaches have to face many challenges. The volume of big data, which requires strong scalability, is a major obstacle, as efficient computational solutions are needed to generate optimal keys for large datasets. Identifying the correct quasi-identifiers is extremely important, as errors can compromise privacy protection [18,19].

Motivation

Big data is related to a large, complex, and increasing datasets that have many, autonomous sources [14]. The emergence of big data technology changed the scenario of information and knowledge-driven systems. In today's world, data has some meaning only if it is discoverable, accessible, sharable, and usable. The more data users share, the more effective and meaningful information systems are. However, when the private and public data is being shared by different people, the integrity of such data needs be preserved.

Problems of information sharing and data privacy:

• Information sharing is a fundamental goal of systems with multiple parties. However, simple data exchange or transmission does not adequately address the issue of privacy.

• Knowing individuals' locations and preferences enables a wide range of useful location-based services. However, public disclosure or leakage of users' location data and movement patterns over time could lead to serious privacy violations.

The main goal of a big data framework is the effective sharing and analysis of data. This process is managed through the implementation of fast data mining algorithms; however, in such systems, data privacy is often compromised. This issue within the big data framework motivates research in the area of privacy preservation for shared data using data mining processes in big data environments.

Literature survey

An overview of the most fascinating and cutting-edge topics in big data and its analytical challenges, including contributions from prominent scientists and recent research papers in the field, is provided below.

Wu et al. [14] have identified and analyzed several critical challenges in big data mining across three key dimensions: data, model, and system levels. Their work emphasizes that to effectively mine big data, the support of high-performance computing platforms is essential. These platforms necessitate systematic designs to fully harness the computational power required to manage big data's scale and complexity.

• In terms of data, the variety of independent information sources and the disparities in the data collection environment frequently result in complex data scenarios, which can include missing values, uncertainties, and inconsistencies. Moreover, challenges such as privacy issues, noise, and data inaccuracies can produce altered or unreliable versions of the dataset. One of the ongoing challenges is the development of safe and secure information-sharing protocols that preserve data integrity and user privacy within the system.

• At the model level, the central challenge is to create a global model by integration of locally discovered

patterns from distributed data sources. This work requires the design of the modern algorithm that is capable of understanding the correlations of models on many sites and combining decisions from these sources. The goal is to get a united, high-quality model reflecting the overall scope of big data and complexity.

• At the system level, the complex, developing relationships between data patterns, models, and data sources should be considered in the big data mining framework. Not only does it handle a large amount of data, but also to install a mechanism to connect unstructured data through a dormant relationship to detect meaningful and predicted patterns. Furthermore, the framework must adapt to temporal dynamics and other influencing factors, ensuring it remains effective as the data landscape evolves.

Overall, the authors regard big data as a major emerging trend, asserting that the demand for effective big data mining solutions is rapidly growing across all domains of science and engineering [14].

Lu et al. [1] note that while classical Homomorphic Encryption-based approaches offer strong cryptographic guarantees, such mechanisms (e.g., Homomorphic Encryption-based protocols using Paillier encryption [15]) may not meet the performance requirements for real-time or large-scale big data mining. In this context, they propose the Personal Computer/Smart Card (PC/SC) protocol to balance computational efficiency with privacy protection when performing similarity computations, including tasks used in clustering, recommendation systems, and anomaly detection. The PC/SC protocol specifically addresses the dual requirements of data confidentiality, by providing a guarantee that sensitive input vectors are kept encrypted or guarded through computation, and computational scalability, to allow for usage within high-volume, distributed big data contexts.

While the authors present a thorough discussion of overall privacy and efficiency issues in big data analytics and present the PC/SC protocol as a potential solution, they point out that more research is required. In particular, more work is needed to solve domain-specific privacy issues that occur in specific applications, such as healthcare, finance, smart cities, and social networks, where context-aware privacy models could be necessary [1].

Ahamad et al. [16] developed a robust privacy protection model in the cloud using AI capabilities, focusing on data sanitization and restoration using the Jaya-based Shark Smell Optimization algorithm. Adopting AI in cloud security increases efficiency and insight in business cloud computing environments. However, current state-of-the-art techniques struggle to achieve optimal privacy, especially for sensitive data, thus requiring separate cleaning and restoration models.

Lekshmy and Abdul Rahiman [17] proposed a method aimed at increasing data privacy through data sanitization, developing privacy-based objective functions, and generating optimal keys using an artificial bee colony algorithm. This approach alleviates data owners' concerns about unauthorized data exposure in the cloud while preserving the utility of the data. However, the model may face limitations when dealing with large datasets in distributed settings.

Li et al. [18] proposed a scheme aimed at increasing the identification of potential fraudulent parties in digital transactions. This scheme eliminates the need for labor-intensive intermediaries in data exchange, ensuring fairness, privacy, and facility through the use of smart contracts, secure transfer protocols, and Ether checks. Despite these efforts to remove third-party mediators, preserving overall fairness and privacy in digital transactions remains a complex challenge.

Shailaja and Guru Rao [19] enhanced data security using privacy-preserving data mining models with trial-based update on whale and particle swarm algorithm (TU-WPA), employing key generation optimization while addressing data sanitization and restoration challenges. This model effectively addresses research problems such as hiding failure rates and saving information. Although the TU-WPA hybrid algorithm is designed to select the optimal key, the complexity of implementing such a hybrid approach can pose challenges in practical applications.

Fang et al. [20] developed a function for an efficient federated learning system in cloud computing that maintains strong privacy protection. The main advantage of this approach is a significant reduction in implementation time and the size of transmitted encrypted data compared to existing secure systems. However, a potential disadvantage is that compromises may be required in certain use cases or specific scenarios to achieve both efficiency and robust privacy preservation.

Challenges

An important challenge is to balance the need to preserve privacy with data utility for classification tasks. Extremely aggressive privacy measures can interfere with the classifier's accuracy. Privacy-protection techniques can make models less meaningful because they complicate the understanding of classification decisions. They may also be time-consuming.

Materials And Methods

The ultimate objective of this research is to develop an optimal key generation approach by creating quasi-identifiers from the data to protect privacy from data providers. In big data, quasi-identifiers can reveal individual identities, leading to attribute disclosure and identity revelation attacks. Therefore, before publishing the data to the service provider, the data is subjected to the identity, which is then subject to the preservation of privacy through the optimum key generation approach. When semi-specialty is known, the data is classified into two categories: sensitive and non-sensitive. Of these, only the sensitive data is retained. This sensitive data is then encrypted to ensure privacy, as illustrated in Figure *2*.

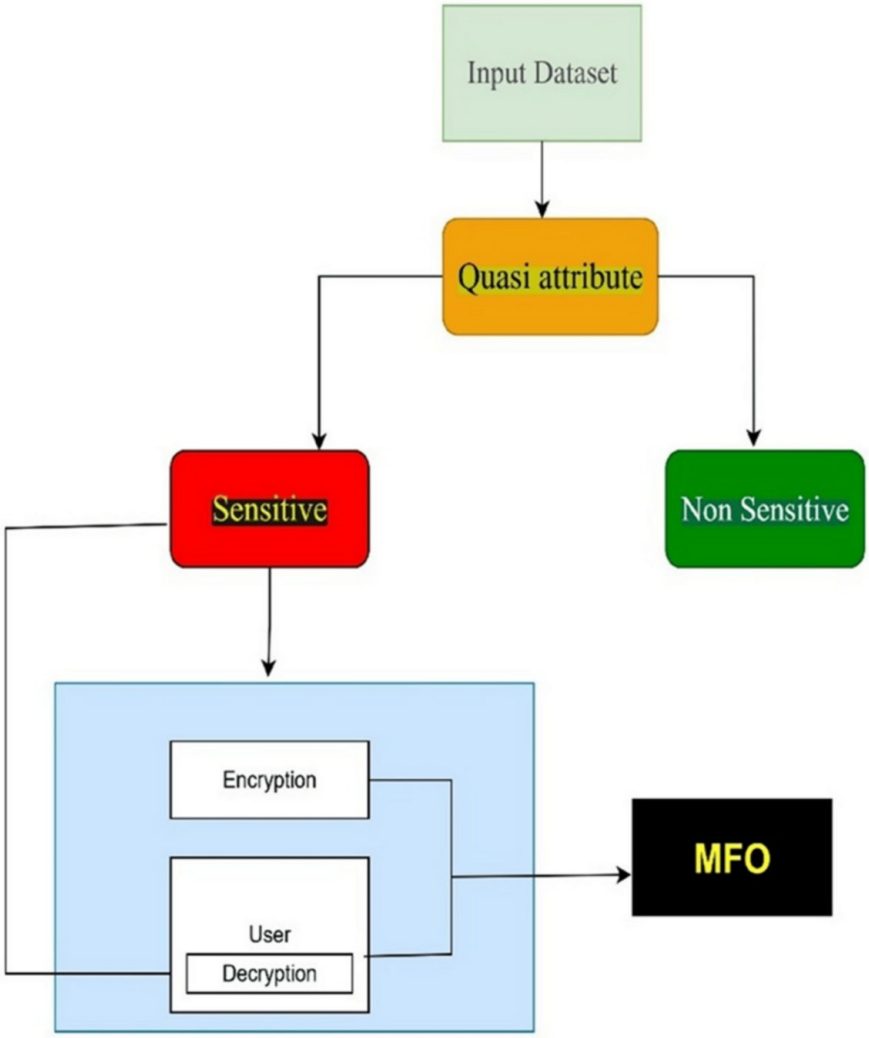

FIGURE 2: Proposed Privacy Preservation Model

MFO, Marine Fish Optimization

Quasi-attribute identification

In this approach, a semi-deduction involves creating a privacy key, dividing data into sensitive and non-sensitive categories, encrypting the sensitive data, and changing data clusters to enhance privacy protection. Data points are assigned values of 0.8 (unique) or 0.5 (frequent).

These unique data points generally pose a higher privacy risk and require special protection, whereas frequent data points have a lower likelihood of disclosing sensitive information and do not constitute a major privacy threat. By considering the impact factor and classifying data as unique or frequent, data sanitization and privacy protection measures can be tailored to the specific characteristics of the data. This approach helps balance data confidentiality, utility, and privacy in the context of big data analysis.

Case 1: $\text{if}\,(\rho < 0.5)\,\&\&\,(A < 1)$

When these conditions are met, the algorithm enters the surrounding phase. Fish can identify and surround a food source. In the context of siege, fish display a coordinated hunting strategy. Initially, each fish searches for food individually, but once a fish locates a food source, they converge and surround it. This collective behavior enhances hunting efficiency, making it difficult for the prey to escape and increasing overall safety for the group. This effective communication among the fish further improves their overall success in capturing fast-moving prey. The concept of siege shows the advantages of cooperative behavior in group hunting. This behavior is represented by the following equation:

$$M^{b+1} = M^{b} + rand * w * M^{best} - \partial * \vec{t} \quad (1)$$

here, $\vec{t}$ denotes the coefficient factor, (w) denotes the weight factor, and (r) denotes the random number.

Case 2: $\text{if}\,(\rho < 0.5)\,\&\&\,(A => 1)$

If this condition is met, the investigation enters the exploration phase. Fish often engage in group behaviors while searching for prey and employ an attractive strategy known as the cyclical stage to locate their target efficiently. During this phase, the fish swim in coordinated patterns or cycles, creating a cyclical movement in the water. This collective movement serves many purposes. First, it allows the fish to focus in a specific area, increasing their chances of finding food. Second, the cyclical movement can confuse or mislead the prey, making it easier to capture. This is a remarkable example of how group behavior and coordinated actions can enhance hunting and fodder detection in the underwater world, helping these aquatic animals to grow in their natural habitat.

$$M^{t+1} = \partial_1' * e^{R_1.R_2}.\cos(2\pi R_2) + M^{best} + M^{b} - M_{sol} \times R_3^{*}\frac{M^{b} - M^{best}}{\partial\,(M^{b} - M^{best})} \quad (2)$$

$$R_1 = \frac{d_g - d_{\min}}{d_{\max} - d_{\min}} \quad (3)$$

here, R_1 denotes the random number, d_g denotes the global distance, $d_{\min}$ denotes the minimal distance, and $d_{\max}$ denotes the maximal distance.

$$R_2 = d_{\max} - (d_{\max} - d_{\min}) \quad (4)$$

$$R_3 = \frac{1}{1 + 1.5.e^{-2.6.r}} e \in (0,1) \quad (5)$$

Results And Discussion

Experimental setup

The experiments were conducted on a system running Windows 10 with 8GB RAM and Python 3.7. Countries' socio-economic profiles contain a wealth of organized information on demographic data, economic activities, employment, health, education, and more. Much of this information can be sensitive or include personally identifiable information in its raw form. This research utilizes these datasets to evaluate the effectiveness of privacy-preservation techniques, such as differential privacy and synthetic data generation, aiming to preserve and protect privacy, while still allowing for useful data.

Experimental analysis

The Marine Fish Optimization (MFO)-Privacy Protection model was used to develop a key generation approach, and its effectiveness was compared with the Marine Predators Privacy Preservation and Jellyfish Privacy Preservation models (Table *1*).

Models	Data privacy	Decryption time (ms)	Encryption time (ms)
Marine Predators Privacy Preservation	1.739	0.02369	0.349
Jelly Fish Privacy Preservation	1.819	0.02399	0.339
MFO-Privacy Preservation	1.869	0.02249	0.329

TABLE 1: Comparative Table

MFO, Marine Fish Optimization

Data privacy comparison

The MFO achieved the highest data privacy score of 1.87, as shown in Figure 3, providing the strongest protection against potential data leakage or estimated attacks. In comparison, the Marine Predators Privacy Preservation model scored the lowest at 1.74, indicating slightly weaker resistance to privacy violations. The Jellyfish Privacy Preservation model performed moderately, achieving a score of 1.82, which is slightly lower than MFO but significantly stronger. These results indicate that the MFO model is the most reliable option for applications where maximum privacy is critical, such as in healthcare and finance.

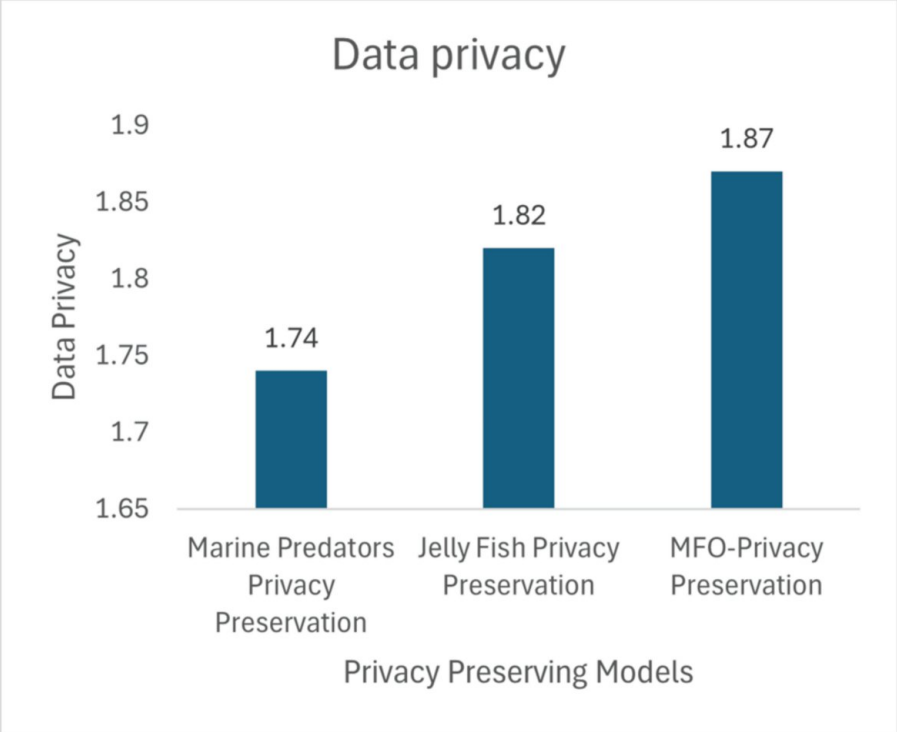

FIGURE 3: Data Privacy

MFO, Marine Fish Optimization

Encryption time comparison

The MFO achieved the fastest encryption time of 0.33ms, as shown in Figure 4, making it the most efficient in data protection. The Jellyfish Privacy Preservation model closely followed with 0.34ms, while the Marine Predators Privacy Preservation model was the slowest at 0.35ms. Although all models have similar encryption performance, MFO shows a slight advantage, making it more suitable for real-time systems.

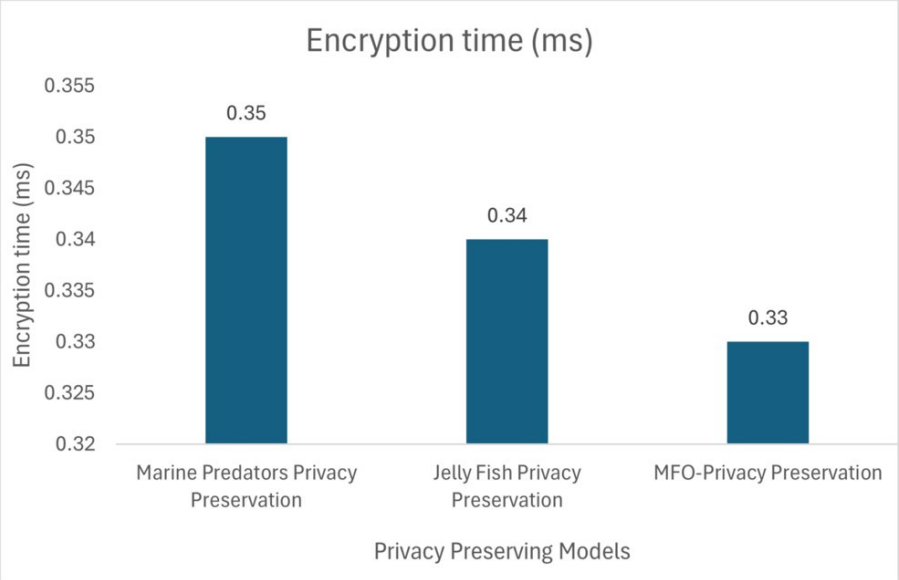

FIGURE 4: Encryption Time

MFO, Marine Fish Optimization

Decryption time comparison

The MFO outperforms the others, achieving the lowest decryption time of 0.0225 ms, as shown in Figure *5*. The Jellyfish is slightly slower at 0.023 ms, while Marine Predators Privacy Preservation records 0.0237 ms. MFO ensures faster access to encrypted data and improves system response, which is crucial for time-sensitive applications.

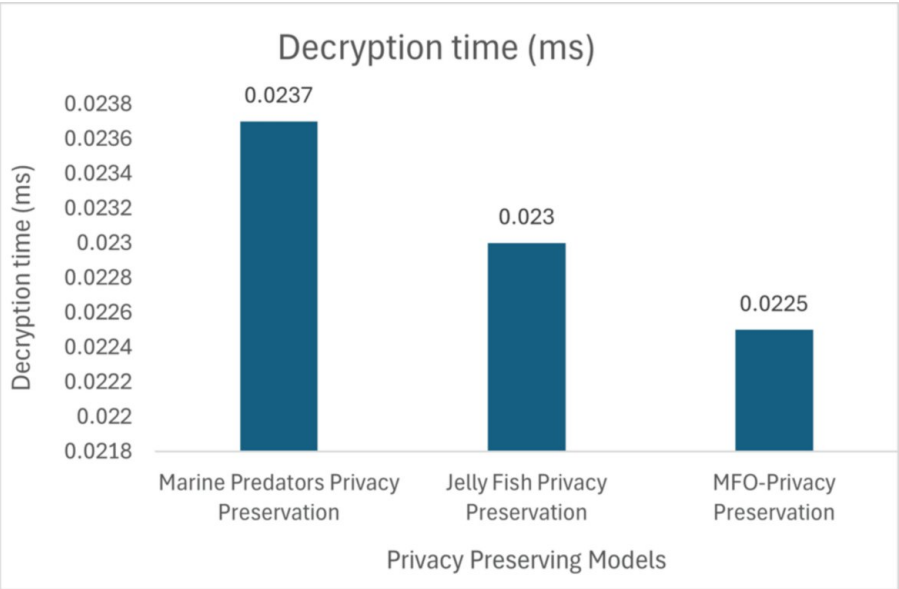

FIGURE 5: Decryption Time

MFO, Marine Fish Optimization

Conclusions

This research addresses these challenges by presenting a new hybrid optimization model designed to enhance the privacy and integrity of data shared in a big data framework. The MFO model uniquely combines the adaptive behaviors of marine predators and jellyfish into a single metaheuristic optimization algorithm. This hybrid approach takes advantage of the exploration capabilities of marine predators and the elasticity and convergence properties of jellyfish, achieving an improved balance

between exploitation and exploration during the optimization process. The proposed MFO models showed excellent performance across major metrics. Specifically, when evaluating Marine Predators Privacy Preservation and Jellyfish Privacy Preservation, the model achieved the highest data privacy score of 1.87, indicating enhanced protection against unauthorized access and potential data violations. It also recorded the fastest encryption and decryption times (0.33 ms and 0.0225 ms, respectively), making it highly suitable for real-time or near-real-time applications in large-scale data environments.

These results validate the effectiveness of the MFO model in optimizing privacy protection while maintaining the computational efficiency required for big data systems. In addition, the model demonstrates promising scalability and adaptability, making it a strong candidate for privacy-integrated big data platforms, cloud-based infrastructures, and secure data-sharing systems. The MFO-Privacy Model also contributes to advancing research aimed at securing big data environments. Overall, this approach provides a viable solution to important issues of privacy, data integrity, and processing performance, paving the way for more secure, efficient, and reliable data-trained circulation in the future.

Appendices

Experimental dataset is available at: https://www.kaggle.com/datasets/nishanthsalian/socioeconomic-country-profiles

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ravi Mante

Acquisition, analysis, or interpretation of data: Ravi Mante, Prashant N. Chatur, Milind B. Waghmare

Drafting of the manuscript: Ravi Mante, Prashant N. Chatur, Milind B. Waghmare

Critical review of the manuscript for important intellectual content: Ravi Mante, Prashant N. Chatur, Milind B. Waghmare

Supervision: Prashant N. Chatur

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Lu R, Zhu H, Liu X, Liu JK, Shao J: Toward efficient and privacy-preserving computing in big data era. IEEE Network. 2014, 28:46-50. 10.1109/mnet.2014.6863131
2. Hu H, Wen Y, Chua TS, Li X: Toward scalable systems for big data analytics: A technology tutorial. IEEE Access. 2014, 2:652-687. 10.1109/access.2014.2332453
3. Li M, Yu S, Ren K, Lou W, Hou YT: Toward privacy-assured and searchable cloud data storage services. IEEE Network. 2013, 27:56-62. 10.1109/mnet.2013.6574666
4. Ahmed R, Karypis G: Algorithms for mining the evolution of conserved relational states in dynamic networks. Knowledge and Information Systems. 2012, 33:603-630. 10.1007/s10115-012-0537-2
5. Lu R, Lin X, Shen X: SPOC: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency. IEEE Transactions on Parallel and Distributed Systems. 2013, 24:614-624. 10.1109/tpds.2012.146
6. Xu L, Jiang C, Wang J, Yuan J, Ren Y: Information security in big data: Privacy and data mining. IEEE Access. 2014, 2:1149-1176. 10.1109/access.2014.2362522
7. Chen R, Sivakumar K, Kargupta H: Collective mining of Bayesian networks from distributed heterogeneous data. Knowledge and Information Systems. 2004, 6:164-187. 10.1007/s10115-003-0107-8
8. Centola D: The spread of behavior in an online social network experiment. Science. 2010, 329:1194-1197. 10.1126/science.1185231
9. Aral S, Walker D: Identifying influential and susceptible members of social networks. Science. 2012, 337:337-341. 10.1126/science.1215842
10. What is big data?. (2012). https://www.ibm.com/think/topics/big-data.
11. Mayer-Schnberger V, Cukier K: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Houghton Mifflin Harcourt, New York; 2014.

12. Twitter Blog. Dispatch from the Denver debate. (2012). Accessed: August 7, 2022: https://www.scirp.org/reference/referencespapers?referenceid=856042.
13. How many public photos are uploaded to Flickr every day, month, year? (6855169886).png. (2012). Accessed: August 12, 2022: https://commons.wikimedia.org/wiki/File:How_many_public_photos_are_uploaded_to_Flickr_every_day,_month,_year%3F_(685
14. Wu X, Zhu X, Wu GQ, Ding W: Data mining with big data. IEEE Transactions on Knowledge and Data Engineering. 2014, 26:97-107. 10.1109/tkde.2013.109
15. Paillier P: Public-key cryptosystems based on composite degree residuosity classes. Advances in Cryptology — EUROCRYPT '99. Stern J (ed): Springer, Berlin, Heidelberg; 1999. 1592:223-238. 10.1007/3-540-48910-X_16
16. Ahamad D, Hameed SA, Akhtar M: A multi-objective privacy preservation model for cloud security using hybrid Jaya-based shark smell optimization. Journal of King Saud University - Computer and Information Sciences. 2022, 34:2343-2358. 10.1016/j.jksuci.2020.10.015
17. Lekshmy PL, Abdul Rahiman M: A sanitization approach for privacy preserving data mining on social distributed environment. Journal of Ambient Intelligence and Humanized Computing. 2019, 11:2761-2777. 10.1007/s12652-019-01335-w
18. Li T, Ren W, Xiang Y: FAPS: A fair, autonomous and privacy-preserving scheme for big data exchange based on oblivious transfer, Ether cheque and smart contracts. Information Sciences. 2021, 544:469-484. 10.1016/j.ins.2020.08.116
19. Shailaja GK, Guru Rao CV: Robust and lossless data privacy preservation: optimal key based data sanitization. Evolutionary Intelligence. 2022, 15:1123-1134. 10.1007/s12065-019-00309-3
20. Fang C, Guo Y, Wang N, Ju A: Highly efficient federated learning with strong privacy preservation in cloud computing. Computers & Security. 2020, 96:101889. 10.1016/j.cose.2020.101889